

Indización en profundidad y aplicación de metadatos a materiales suplementarios en la edición de revistas

Luis Rodríguez-Yunta

23 octubre 2013

Rodríguez-Yunta, Luis (2024). "Indización en profundidad y aplicación de metadatos a materiales suplementarios en la edición de revistas". *Anuario ThinkEPI*, v. 8, pp. 207-210.



Resumen: Entre las buenas prácticas de la edición de revistas científicas se llama la atención sobre la recomendación publicada por la *NISO* referente a la inclusión de materiales suplementarios. Se justifica la necesidad de dar un tratamiento específico a estos materiales, que deben contar con metadatos concretos. Se relacionan estas prácticas con la tendencia de indización en profundidad practicada por productores de bases de datos documentales, como *ProQuest*.

Palabras clave: Revistas científicas, Material gráfico, Bases de datos, Indización en profundidad, Metadatos.

Title: Deep indexing and the application of metadata to supplemental materials in journal publishing

Abstract: Among the best practices of academic journals publishing, this note focusses on the *NISO* recommendation about the inclusion of supplemental materials. The paper justifies the need to give special treatment to these materials, which require specific metadata. Finally it relates these practices with the deep indexing trend among bibliographic database producers such as *ProQuest*.

Keywords: Academic journals, Graphic material, Databases, Deep indexing, Metadata.

1. El artículo no es la única unidad de información en una revista científica digital

Hace más de diez años, **Elena Fernández-Sánchez** e **Isabel Fernández-Morales** (2000), documentalistas del ya desaparecido *Cindoc*, describían los procesos técnicos de la edición de revistas en internet. Entre las ventajas de la elaboración de publicaciones electrónicas puras frente a las versiones digitalizadas a partir de una impresa se citaba:

- tratamiento de la parte gráfica más adaptada al nuevo medio;
- espacio visual más rico y colorista, con opción de aumentar imágenes a voluntad del usuario;
- posibilidad de incorporar otros soportes como vídeos, audios o realidad virtual.

En este artículo se decía: "En el caso de los diarios se observa que las versiones digitales son meras copias de las realizadas en papel. Por ejemplo, no se explota el concepto de hipertextualidad de una noticia a otra, ni tampoco existe una organización visual adaptada al formato de pantalla".

Esta afirmación no puede mantenerse hoy en

día para la prensa digital que ya ha dado el salto hace varios años para adaptarse a las demandas de los lectores en internet. Sin embargo, este salto aún está pendiente para las numerosas revistas científicas que se limitan a colgar textos en pdf manteniendo una imagen exacta a la versión impresa, incluso en publicaciones que carecen de ella.

La edición de revistas científicas en muchos casos aún tiene pendiente abordar la revolución digital con todas sus consecuencias. Hace falta "pensar en digital", abandonar toda dependencia de la versión impresa, que puede mantenerse si es necesario como estrategia de distribución, pero que no debe frenar la explotación y difusión a través de internet. En este nuevo medio los usuarios pueden acceder a los artículos directamente a partir de un buscador o una base de datos sin pasar por las páginas de presentación y sumario de la revista. Pero los accesos a través de los buscadores van más allá, también pueden dar entrada a cualquier elemento que constituye un fichero independiente (imágenes en jpg o tiff, un fichero *Excel* de datos, un anexo en *Word*,...). El artículo no es la única unidad de información objeto de una posible recuperación documental.

2. Datos de identificación de los elementos suplementarios de los artículos

Un aspecto sintomático a tener en cuenta para valorar la adaptación de una edición a la era digital consiste en analizar cómo se trata el material gráfico que acompaña a los artículos de una revista. Estos elementos deben poder identificarse y ser objeto de recuperación de forma independiente y con cierta pertinencia. Los ficheros anejos, y también las imágenes, gráficos o tablas precisan de datos de identificación propios que pueden ser diferentes o complementarios de los del artículo en el que se ubican:

- Pueden ser de autoría diferente a la del texto. Es necesario identificar créditos para figuras, ilustraciones o fotografías, así como sobre los datos de donde proceden los gráficos o tablas.
- Pueden estar sujetos a una regulación de derechos independiente a la del texto.
- Deben contar con título y/o descripción precisos. El título puede enriquecerse con una leyenda, que aporta información textual referida exclusivamente a esta parte del documento.
- En la edición puede proceder de un fichero independiente, con características técnicas determinadas (soporte, formato, resolución,...).

“El artículo no es la única unidad de información objeto de una posible recuperación documental”

La cuestión de los derechos de autor puede justificar por sí sola la necesidad de un tratamiento específico de los materiales adicionales de un artículo. Por ejemplo, un autor puede autorizar que se reproduzcan partes de su texto siempre que sea citado, pero no que se haga lo mismo con el material gráfico.

En relación con las imágenes o fotografías, éstas pueden haber sido adquiridas a terceros. El hecho de que se incorporen a un artículo no significa que le afecten las mismas condiciones de *copyright* o *copyleft*. En cualquier caso, se trata de una información relevante que debe especificarse



Artículo en Science Direct con visualización de figuras:
<http://www.sciencedirect.com/science/article/pii/S1135307410700296>

en los créditos de cada figura.

En el campo de la imagen científica, **Anne-Vinciane Doucet** (2008) afirma: “es característico que la imagen científica se acompañe de un comentario que complete la información visual. Se recomienda que el texto aporte datos sobre la técnica fotográfica utilizada y la tasa de ampliación, para no confundir al espectador con respecto a la escala de lo que observa”.

Aun cuando la autora, que realizó una tesis doctoral sobre este tema en la *Universidad de Granada*, aboga por la normalización en la descripción y palabras clave dentro de un banco de imágenes, sus apreciaciones pueden aplicarse perfectamente a la inclusión de estas mismas imágenes en las revistas científicas.

En los últimos años se está promoviendo la difusión de los ficheros de datos dentro de los repositorios (**Hernández-Pérez; García-Moreno, 2013**). El mismo interés puede tener poder aislar los gráficos y tablas estadísticas que los presentan dentro de un texto. Para **David Shotton** (2009; 2012), la edición semántica es la revolución en marcha en las publicaciones científicas. Las cinco claves de su propuesta son:

- asegurar el *peer review*;
- distribuir en acceso abierto;
- enriquecimiento de contenidos con toda la potencialidad de la web semántica;
- accesibilidad de los datos;
- representación de metadatos legibles por máquinas, tanto para el artículo como para las referencias citadas.

En definitiva, lo que se promueve es un mejor aprovechamiento de las posibilidades que ofrece realmente la era digital. Ello implica aspectos concretos como utilizar html o xml además de o

en lugar del pdf, incrementar el uso de DOIs para todas las fuentes útiles de información y también la indización específica de los materiales suplementarios. Sin duda, trabajos que apuntan hacia la necesidad de profesionalizar las tareas técnicas en la edición de revistas o en las plataformas editoriales (Rodríguez-Yunta; Tejada-Artigas, 2013).

“La Niso considera materiales suplementarios los elementos integrales o esenciales para la lectura del texto y los contenidos adicionales que pretenden enriquecerlo”

3. Metadatos para los materiales suplementarios: recomendaciones Niso

Los editores de revistas ofrecen en muchos casos un tratamiento diferenciado para las ilustraciones, figuras o materiales anexos a un artículo. Un ejemplo de resalte de las figuras en la edición puede verse en las revistas de *Elsevier* a través de *Science Direct*. En la previsualización de un artículo, tras el resumen y las palabras clave, se muestran las figuras con sus leyendas, en formato reducido y a veces meramente esquemático para las tablas. Igualmente las revistas de la plataforma de acceso abierto *revues.org*, muestran las ilustraciones de forma independiente al final del documento con sus propios créditos y un url para citación o acceso directo. Un modelo más desarrollado es el que utilizan las publicaciones de *PLoS*. En la cabecera de cada artículo se presentan visualizaciones reducidas de las figuras; cada una de ellas puede verse en mayor tamaño, exportar en diferentes formatos (ppt, png o tiff), y cuentan con resumen y DOI independientes.

En enero de 2013, la *National Information Standards Organization (Niso)* ha publicado un documento sobre buenas prácticas para el tratamiento de los materiales suplementarios en artículos de revistas (*Niso RP-15-2013*), que pretende aportar unas pautas para la mejora y normalización de la tarea de los editores en este aspecto. El documento aclara el alcance del concepto de materiales suplementarios, en el que se incluyen tanto elementos integrales o esenciales para la lectura del texto (como tablas, figuras o ficheros de datos necesarios para la demostración de ideas o investigaciones), como los contenidos adicionales que pretenden enriquecerlo (una bibliografía extendida o cualquier representación o aplicación que suponga una extensión de las ideas del artículo).

Entre las recomendaciones se especifican tres áreas que deben tenerse en cuenta para mantener la necesaria consistencia en la presentación y accesibilidad de estos elementos:

- sumario del número: donde debe mencionarse la presencia de material suplementario.
- visualización dentro del artículo: los enlaces a materiales adicionales deben situarse al principio de la primera página o en la cabecera del artículo en la edición web.
- visualización de los propios materiales suplementarios de forma independiente.

El documento *Niso RP-15-2013* incluye una especificación de los metadatos aplicables a la descripción de materiales suplementarios en los artículos de revistas. Establece también unos elementos mínimos recomendables:

- identificador persistente para cada material suplementario (por ejemplo un DOI);
- identificador persistente para el artículo (por ejemplo un DOI);
- tipo de relación entre el material suplementario y el artículo (integral, adicional, o ambas);
- nombre del fichero o enlace externo al mismo;
- formato del fichero del material suplementario.

“Los investigadores valoran positivamente la posibilidad de realizar búsquedas sobre materiales desagregados de los artículos, en particular tablas y figuras”

4. Implicación para las bases de datos documentales: indización en profundidad

La *Niso RP-15-2013* también indica la conveniencia de que los servicios de indización y resumen recojan la información sobre los materiales adicionales. La responsabilidad del editor es asegurar que estos materiales sean recuperables y fáciles de descubrir (*ensuring discoverability and findability*). Ello implica tanto la presencia en buscadores como en las bases de datos documentales.

En este sentido, el documento reconoce la función esencial y vigencia de estos productos: “La cobertura de los resúmenes y de la indexación es esencial para un descubrimiento óptimo. Incluso en estos tiempos de múltiples servicios de descubrimiento, los académicos continúan basándose en los resúmenes y en la indexación para asegurar que su búsqueda sea exhaustiva” (*Niso*, 2013, p. 8).

Entre los productores de servicios de información, la posibilidad de descubrir materiales adicionales dentro de los registros de una referencia se ofrece ya en algunas bases de datos documentales. Un ejemplo de ello es el servicio *Deep indexing* (indización en profundidad), que permite la búsqueda de figuras, tablas, gráficos, mapas, diagramas y otras ilustraciones indizadas a partir de los artículos de revistas científicas. Esta herramienta fue desarrollada inicialmente dentro de la plataforma *CSA illustrata* en 2005-2006 (**Tenopir; Sandusky; Casado**, 2006) y se pudo utilizar en línea por los usuarios desde 2007. Aunque *CSA illustrata* ha desaparecido al ser adquirida por el grupo *ProQuest*, la herramienta permanece dentro de la plataforma de búsqueda de este productor y afecta a bases de datos como *GeoRef*, *Agricola* o *Aquatic sciences*.

Un estudio realizado por **Robert J. Sandusky** y **Carol Tenopir** (2008) muestra que los investigadores valoran muy positivamente la posibilidad de realizar búsquedas precisas sobre estos materiales desagregados de los artículos, en particular tablas y figuras.

Esta indización en profundidad deberá aplicarse en todos los productos relacionados con el acceso a artículos de revistas científicas, en la medida en la que los editores faciliten datos de identificación de forma rigurosa y clara para todos estos elementos suplementarios que suponen una demanda específica para los usuarios de documentación científica. Si solamente los buscadores ofrecen estos servicios, las bases de datos documentales habrán perdido otra batalla por mantener a sus usuarios.

Bibliografía

Doucet, Anne-Vinciane (2008). "El análisis de la normalización de los elementos de contenido de leyendas y palabras clave en fotografías científicas". *Acimed*, v. 18, n. 2. <http://digibug.ugr.es/handle/10481/21602>

Fernández-Sánchez, Elena; Fernández-Morales, Isabel (2000). "Consideraciones sobre la edición electrónica de revistas en internet". *El profesional de la información*, v. 9, n. 3, pp. 4-12. <http://eprints.rclis.org/13229>



Artículo de editorial *PLOS* con visualización de figuras:

<http://www.plosbiology.org/article/info%3Adoi%2F10.1371%2Fjournal.pbio.1001672>

Hernández-Pérez, Tony; García-Moreno, María-Antonia (2013). "Datos abiertos y repositorios de datos: nuevo reto para los bibliotecarios". *El profesional de la información*, v. 22, n. 3, pp. 259-263. <http://dx.doi.org/10.3145/epi.2013.may.10>

Niso RP-15-2013, Recommended practices for online supplemental journal article materials. <http://www.niso.org/publications/rp/rp-15-2013>

ProQuest Deep Indexing. http://www.csa.com/le_products/pqdeepindex.php

Rodríguez-Yunta, Luis; Tejada-Artigas, Carlos-Miguel (2013). "El editor técnico: un perfil necesario para la profesionalización de la edición de revistas científicas en el entorno digital". *Anales de documentación*, v. 16, n. 2. <http://eprints.rclis.org/20712/>

Sandusky, Robert J.; Tenopir, Carol (2008). "Finding and using journal-article components: Impacts of disaggregation on teaching and research practice". *Journal of the American Society for Information Science and Technology*, v. 59, n. 6, pp. 970-982. <http://dx.doi.org/10.1002/asi.20804>

Shotton, David (2009). "Semantic publishing: the coming revolution in scientific journal publishing". *Learned publishing*, v. 22, n. 2, pp. 85-94. <http://dx.doi.org/10.1087/2009202>

Shotton, David (2012). "The five stars of online journal articles — a framework for article evaluation". *D-Lib magazine*, v. 18, n. 1/2. <http://dx.doi.org/10.1045/january2012-shotton>

Tenopir, Carol; Sandusky, Robert J.; Casado, Margaret (2006). "The value of CSA deep indexing for researchers (executive summary)". *School of Information Sciences Publications and Other Works*. <http://goo.gl/rJ8aCk>